Biost 518
Applied Biostatistics II

Scott S. Emerson, M.D., Ph.D.
Professor of Biostatistics
University of Washington

Lecture 8:
Multiple Regression:
Effect Modification

February 4, 2008

1

Lecture Outline

- Effect modification
- Examples
  - Administrative duties modifying the association between salary and sex
  - SEP by height, age, sex

2

Effect Modification

3

Effect Modifier

- The association between Response and POI differs in strata defined by effect modifier
    - Statistical term: "Interaction"
  - Depends on the measurement of effect
    - Summary measure
      - Mean, geometric mean, median, proportion, odds, hazard, etc.
    - Comparison across groups
      - Difference, ratio

4

Analysis of Effect Modification

- When the scientific question involves effect modification, analyses must be within each stratum separately
  - If we want to estimate degree of effect modification or test for its existence:
    - A regression model will typically include
      - Predictor of interest
      - Effect modifier
      - A covariate modeling the interaction (usually product)

5

Model for Effect Modification

- Typical model for effect modification
  - Include "main effects" (can be bad not to)
    - *X* (or predictors that involve only *X*)
    - *W* (or predictors that involve only *W*)
  - Include "interactions"
    - Predictor(s) derived from both *X* and *W*

$$g[\theta \mid X_i, W_i] = \beta_0 + \beta_X \times X_i + \beta_W \times W_i + \beta_{XW} \times (XW)_i$$
$$= \beta_0 + \beta_X \times X_i + \beta_W \times W_i + \beta_{XW} \times X_i \times W_i$$

6

Interpretation of Parameters

$$g[\theta \mid X_i, W_i] = \beta_0 + \beta_X \times X_i + \beta_W \times W_i + \beta_{XW} \times X_i \times W_i$$

- Usual approach a bit more difficult
  - We can try using the idea of "comparison of θ across groups differing by 1 unit in corresponding predictor but agreeing in other modeled predictors"
  - However, terms involving two scientific variables makes this approach difficult

7

Intercept

$$g[\theta \mid X_i, W_i] = \beta_0 + \beta_X \times X_i + \beta_W \times W_i + \beta_{XW} \times X_i \times W_i$$

- Interpretation of intercept straightforward
  - β_0 corresponds to *X*= 0, *W*= 0
    - May not be scientifically meaningful

8

Slopes for Main Effects

$$g[\theta \mid X_i, W_i] = \beta_0 + \beta_X \times X_i + \beta_W \times W_i + \beta_{XW} \times X_i \times W_i$$

- Interpretation of main effects
  - β_X corresponds to 1 unit difference in *X* holding *W* and *(X×W)* constant
    - So 1 unit difference in *X* when *W*= 0
    - May not be scientifically meaningful
  - β_W corresponds to 1 unit difference in *W* holding *X* and *(X×W)* constant
    - So 1 unit difference in *W* when *X*= 0
    - May not be scientifically meaningful

9

Slope for interaction

$$g[\theta \mid X_i, W_i] = \beta_0 + \beta_X \times X_i + \beta_W \times W_i + \beta_{XW} \times X_i \times W_i$$

- Interpretation of interaction difficult
  - β_{XW} corresponds to 1 unit difference in *(X×W)* holding *X* and *W* constant
    - Impossible, so we need another way to interpret this slope parameter

10

Consider Scientific Predictors

$$g[\theta \mid X_i, w] = \beta_0 + \beta_X \times X_i + \beta_W \times w + \beta_{XW} \times X_i \times w$$
$$= (\beta_0 + \beta_W \times w) + (\beta_X + \beta_{XW} \times w) \times X_i$$

In stratum with $W = w$

Intercept : $(\beta_0 + \beta_W \times w)$ corresponds to $X_i = 0$

Slope : $(\beta_X + \beta_{XW} \times w)$ compares groups differing by 1 unit in X

β_{XW} is difference in X slope per 1 unit difference in W

11

Consider Scientific Predictors

$$g[\theta \mid x, W_i] = \beta_0 + \beta_X \times x + \beta_W \times W_i + \beta_{XW} \times x \times W_i$$
$$= (\beta_0 + \beta_X \times x) + (\beta_W + \beta_{XW} \times x) \times W_i$$

In stratum with $X = x$

Intercept : $(\beta_0 + \beta_X \times x)$ corresponds to $W_i = 0$

Slope : $(\beta_W + \beta_{XW} \times x)$ compares groups differing by 1 unit in W

β_{XW} is difference in W slope per 1 unit difference in X

12

Symmetry of Effect Modification

- Note that if X modifies the association between Y and W, then W modifies the association between Y and X
  - Aside: Confounding need not be symmetric
    - W can confound the association between Y and X, but X not confound the association between Y and W
      - W and X associated in the sample
      - Y and X not associated after adjusting for W
      - Y and W associated after adjusting for X

13

Inference for Effect Modification

$$g[\theta \mid X_i, W_i] = \beta_0 + \beta_X \times X_i + \beta_W \times W_i + \beta_{XW} \times X_i \times W_i$$

- No effect modification if $\beta_{XW} = 0$
  - Hence, inference about existence of effect modification tests that $\beta_{XW} = 0$
    - We can perform such inference using standard regression output for the corresponding slope parameter

14

Inference for Main Effect Slope

$$g[\theta \mid X_i, W_i] = \beta_0 + \beta_X \times X_i + \beta_W \times W_i + \beta_{XW} \times X_i \times W_i$$

- Interpretation of $\beta_X = 0$
  - Same intercept in all strata defined by *W*
  - Generally a <u>very</u> uninteresting question
  - We rarely make inference on main effect slopes by themselves

15

Inference About Effect of X

$$g[\theta \mid X_i, W_i] = \beta_0 + \beta_X \times X_i + \beta_W \times W_i + \beta_{XW} \times X_i \times W_i$$

- Response parameter not associated with *X* if $\beta_X = 0$ <u>AND</u> $\beta_{XW} = 0$
  - We will need to construct special tests that both parameters are simultaneously 0
    - The t tests given in regression output consider only one slope parameter at a time

16

Stata: Testing Multiple Slopes

- Stata has easy method for performing test that multiple parameters are simultaneously 0
  - Perform any regression command
  - Then use "`test var1 var2 …`"
    - Provides P value of the hypothesis test based on most recently executed regression command *of any type of regression*

17

Ex: Salary by Sex and Admin

- Does sex modify the association between mean salary and administrative duties
  - With two binary variables, modeling interaction by product is the obvious choice

$$E[Sal \mid Fem, Adm] = \beta_0 + \beta_A \times Adm_i + \beta_F \times Fem_i + \beta_{AF} \times Adm_i \times Fem_i$$

18

Ex: Stata output

```
. g admfem= admin * female
. regress salary admin female admfem if year==95,
Linear regression          Number of obs =    1597
                           F(  3,  1593) =  125.26
                           Prob > F      =  0.0000
                           R-squared     =  0.1615
                           Root MSE      =  1866.9
       |           Robust
salary |    Coef.  StdErr  t     P>|t|   [95% CI]
 admin |  1951.378   176  11.06 0.000  1605   2297
female | -1226.234    95 -12.86 0.000 -1413  -1039
admfem | -461.9072   342  -1.35 0.177 -1132    208
 _cons |  6506.607    62 105.25 0.000  6385   6627
```

19

Ex: Descriptive Statistics

- Note that with two binary variables, the regression parameters agree exactly with the corresponding group sample means

```
. table admin female if year==95, co(mean salary)

         |       female
   admin |     Male    Female
Nonadmin | 6506.607  5280.373
   Admin | 8457.985  6769.844
```

20

Ex: Inference About Eff Mod

- Does sex modify association between mean salary and administrative duties?
  - Estimate that the "administrative supplement" averages $462 less for women than men
    - 95% CI: $1132 less to $208 more
    - Not statistically significant: P = 0.177

```
       |              Robust
salary |    Coef.  StdErr  t     P>|t|   [95% CI]
 admin |  1951.378   176  11.06 0.000  1605  2297
female | -1226.234    95 -12.86 0.000 -1413 -1039
admfem | -461.9072   342  -1.35 0.177 -1132   208
 _cons |  6506.607    62 105.25 0.000  6385  6627
```

21

Ex: Inference About Sex Assoc

- Is sex associated with mean salary?
  - Need to test that slope parameters for `female` and `admfem` are simultaneously 0

```
. test female admfem
 ( 1)  female = 0
 ( 2)  admfem = 0

       F(  2,  1593) =   95.90
            Prob > F =    0.0000
```

22

Ex: Inference for Admin Assoc

- Are administrative duties associated with mean salary?
  - Need to test that slope parameters for `admin` and `admfem` are simultaneously 0

```
. test admin admfem
 ( 1)  admin = 0
 ( 2)  admfem = 0

       F(  2,  1593) =   74.15
            Prob > F =    0.0000
```

23

Continuous Predictors

- Modeling interactions with continuous predictors is conceptually more complicated
  - Is a multiplicative interaction at all a reasonable model for the data?
  - Nonetheless, this is the most common way we detect interactions
    - I would caution against using the model as predictions without carefully examining the data
      - But this can be difficult, too

24

Example: SEP "Normal Ranges"

- We want to find normal ranges for somatosensory evoked potential (SEP)
  - As a first step, we want to consider important predictors of nerve conduction times
    - If any variables such as sex, age, height, race, etc. are important predictors of nerve conduction times, then it would make most sense to obtain normal ranges within such groups

25

Example: SEP "Normal Ranges"

- Scientifically, we might expect that height, age, and sex are related to the nerve conduction time
  - Nerve length should matter, and height is a surrogate for nerve length
  - Age might affect nerve conduction times: People slow down with age
  - Sex: Men are <u>SO</u> fragile

26

Example: SEP "Normal Ranges"

- Prior to looking at the data, we can also consider the possibility that interactions between these variables might be important
  - Height - age interaction?
    - Do we expect the difference in conduction times between 6 foot tall and 5 foot tall 20 year olds to be the same as the difference in conduction times between 6 foot tall and 5 foot tall 80 year olds?

27

Example: SEP "Normal Ranges"

- We might suspect such an interaction due to the fact that height may not be as good a surrogate for nerve length in older people
  - With age, some people tend to shrink due to osteoporosis and compression of intervertebral discs
    - It is not clear that nerve length would be altered in such a process

28

Example: SEP "Normal Ranges"

- Thus, in young people, differences in height probably are a better measure of nerve length than in old people
    - Tall old people probably have been tall always
    - Short old people will include some who were much taller when they were young

29

Example: SEP "Normal Ranges"

- We can also consider the possibility of three way interactions between height, age, and sex
    - Osteoporosis affects women far more than men
        - Hence, we might expect the height - age interaction to be greatest in women and not so important in men

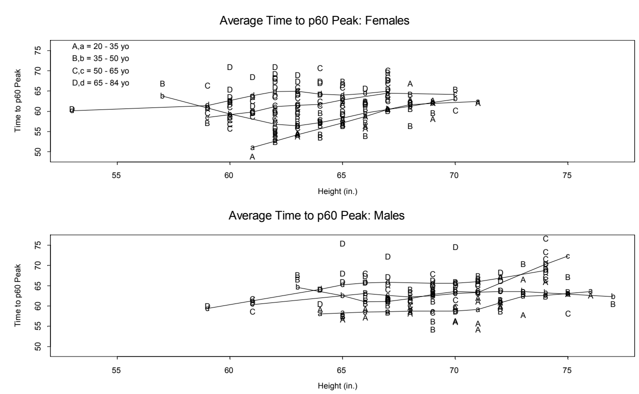
30

Example: SEP "Normal Ranges"

- A two way interaction between height and age that is different between men and women defines a three way interaction between height, age, and sex

31

Stratified Scatterplots



32

Example: SEP "Normal Ranges"

- Defining a regression model with interactions
  - We must create variables to model the three way interaction term

33

Example: SEP "Normal Ranges"

- Furthermore, it is a VERY GOOD idea to include all "main effects" and "lower order interactions" in the model as well
  - "main effects": the individual variables which contribute to the interaction
  - "lower order terms": all interactions that involve some combination of the variables which contribute to the interaction

34

Example: SEP "Normal Ranges"

- Most often, we lack sufficient information to be able to guess what the true form of an interaction might be
  - The most popular approach is thus to consider multiplicative interactions
    - Create a new variable by merely multiplying the two (or more) interacting predictors

35

Example: SEP "Normal Ranges"

- Thus for this problem we could create variables
  - HA = Height * Age
  - HM = Height * Male
  - AM = Age * Male
  - HAM = Height * Age * Male

36

Example: SEP "Normal Ranges"

- Interpretation of the model parameters
  - In the presence of higher order terms (powers, interactions) interpretation of parameters is not easy
    - We can no longer use "the change associated with a 1 unit difference in predictor holding other variables constant"
      - It is generally impossible to hold other variables constant when changing a covariate involved in an interaction
      - If not impossible, it is often uninteresting

37

Example: SEP "Normal Ranges"

Interpretation of the model in terms of the SEP height relationship within age-sex strata

38

Example: SEP "Normal Ranges"

$$E(p60 \mid Ht, Age, Male) = \beta_0 + \beta_H Ht + \beta_A Age + \beta_M Male$$
$$+ \beta_{HA} HA + \beta_{HM} HM + \beta_{AM} AM + \beta_{HAM} HAM$$

p60 - Height relationship for Age = a :

Sex	Intercept	Slope
F	$(\beta_0 + \beta_A a)$	$(\beta_1 + \beta_{HA} a)$
M	$(\beta_0 + \beta_M + (\beta_A + \beta_{AM})a)$	$(\beta_H + \beta_{HM} + (\beta_{HA} + \beta_{HAM})a)$

39

Example: SEP "Normal Ranges"

- From the above, we see the importance of including the main effects and lower order terms
  - E.g., leaving out the height - sex interaction is tantamount to claiming that the p60 - height relationship among newborns is the same for the two sexes
    - (It might be, but the chance that our lines would predict the truth is very slight-- we are trying to approximate relationships in other age ranges)

40

Example: Regression Output

```
. regress p60 height age male HA HM AM HAM
```

p60	Coef	SE	t	P>|t|	[95% CI]	
height	1.38	.363	3.81	0.000	.666	2.09
age	1.13	.425	2.66	0.008	.292	1.97
male	75.0	32.3	2.32	0.021	11.3	138
HA	-.015	.007	-2.26	0.025	-.028	-.0019
HM	-1.12	.483	-2.34	0.020	-2.08	-.176
AM	-1.16	.582	-2.00	0.047	-2.31	-.0170
HAM	.0175	.009	2.00	0.047	.0002	.0347
_cons	-36.4	23.5	-1.55	0.122	-82.7	9.82

Aside: Subgroup Analysis

- If I restrict analysis to females, estimates are the same in this "saturated" model
  - (Restricting by age or height would differ due to "borrowing information across groups)
  - Inference can differ due to the estimate of the residual standard error

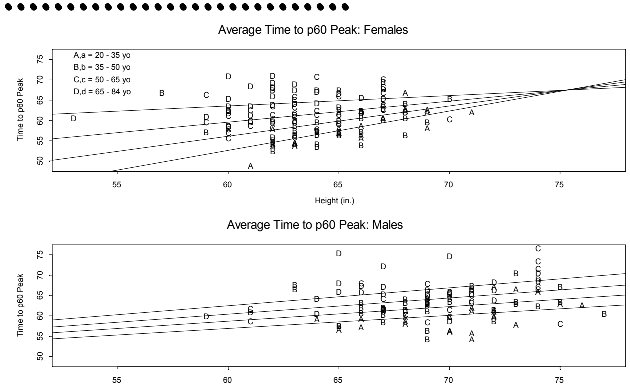
```
. regress p60 height age HA   if male==0
```

p60	Coef	SE	t	P>|t|	[95% CI]	
height	1.38	.361	3.82	0.000	.665	2.10
age	1.13	.424	2.67	0.009	.292	1.97
HA	-.015	.007	-2.27	0.025	-.028	-.002
_cons	-36.4	23.4	-1.56	0.122	-82.7	9.86

Interpreting Estimates

- Figuring out what all these estimates mean is nearly impossible
  - I find it easiest to graph the predicted values

Lines Predicted By Model



Example: SEP "Normal Ranges"

- From the inference, we find a statistically significant three way interaction
  - P= .0347
- This would argue that I should make predictions based on a model including the 3-way interaction
  - But…

45

Influence of Individual Cases

- I always worry that interactions might be significant only because of a single "outlier"
  - If that were the case, I might choose not to include the interaction (but I always include the case)
  - Looking ahead: I can "diagnose" such a problem by assessing the influence of each case

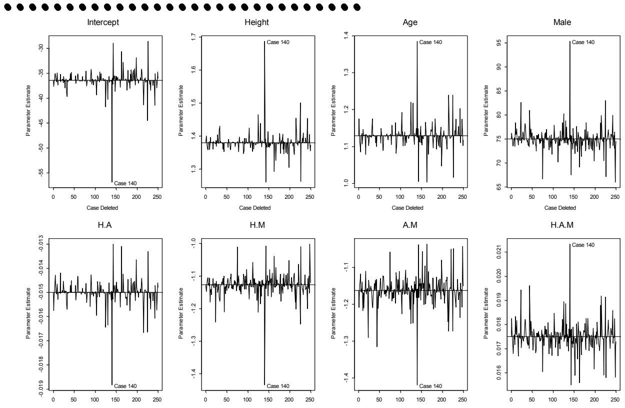
46

Example: SEP "Normal Ranges"

- I am now interested in ensuring that the evidence for an interaction is not based solely on a single person's observation
  - Hence, I consider 250 different regressions in which I leave out each case in turn
  - I plot the slope estimates and P values for each variable as a function of which case I left out
    - Case 0 corresponds to using the full data set

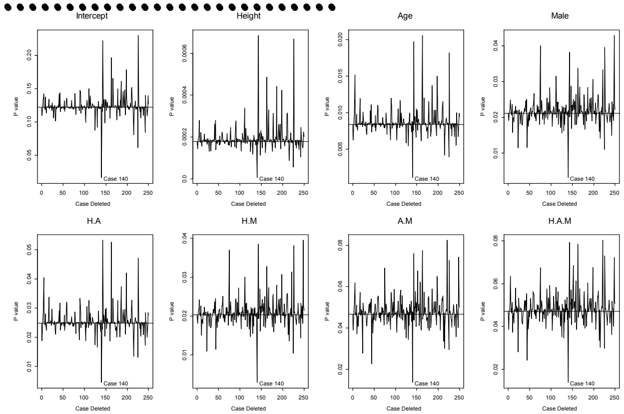
47

Influence on Estimates



48

Influence on P values



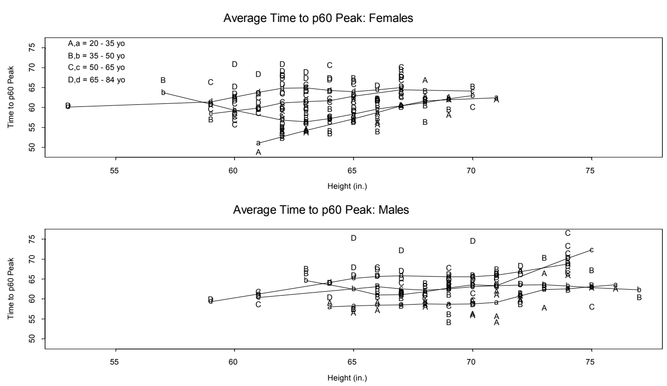
49

Example: SEP "Normal Ranges"

- Contrary to what I was afraid of, the only influential case actually lessened the evidence of an interaction
  - When Case 140 is removed from the data, the evidence for an interaction is a larger estimate and a lower P value
  - We can examine the scatterplot to see why Case 140 might be so influential

50

Stratified Scatterplots



51

Example: SEP "Normal Ranges"

- So now what do I do with Case 140
  - From the influence diagnostics, I now feel comfortable with the fact that the data really do suggest a three way interaction

52

Example: SEP "Normal Ranges"

- Personally, I do NOT remove the case from the dataset when making my prediction intervals
  - I do not know why Case 140 is so unusual
  - It is possible that people like her are actually more prevalent in the population than my sample would suggest
    - My best guess is that she represents 0.4% of the population, so leave her in

53